B O N N E V I L L E   P O W E R   A D M I N I S T R A T I O N

Synchrophasor Data Mining

Tony Faris

i-PCGRID

March 2015

Scope

- Number of data frames reported by a single PMU:

Time Period	Number of Data Frames (30 fps)	Number of Data Frames (60 fps)
1 second	30	60
1 minute	1800	3600
1 hour	108,000	216,000
1 day	2.592 million	5.184 million
30 days	77.76 million	155.52 million
365 days	946.08 million	1.892 billion

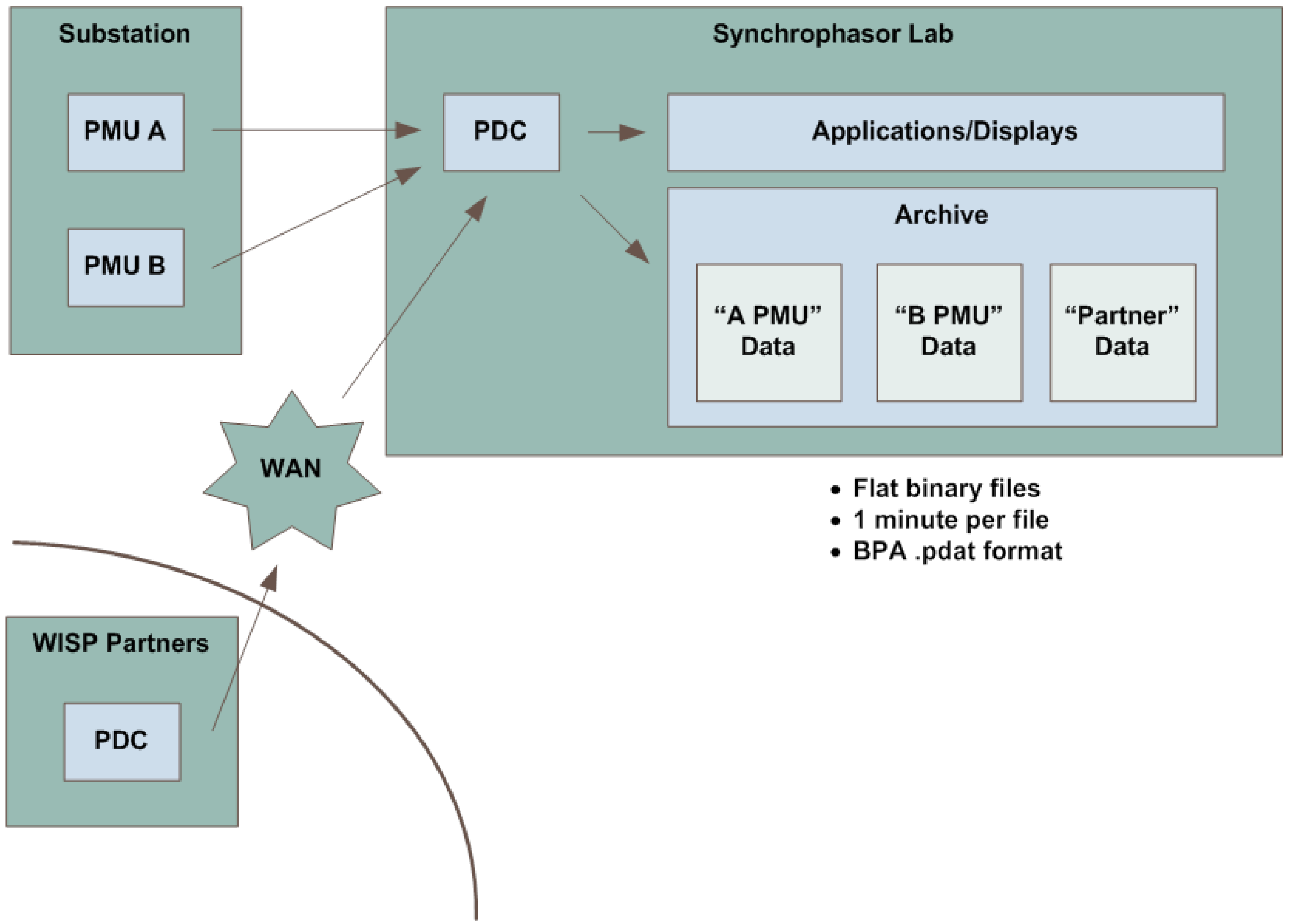
Scope

- For a "fully-loaded" PMU data frame at BPA:

Data Type	Number of Data Points
Phasor Magnitude	14
Phasor Angle	14
Analog (MW and MVAR)	12
Digital Bits	16 (1 digital word)
Frequency	1
Rate of Change of Freq.	1
Status Flags	4
TOTAL	**62**

- One year = 177.3 billion points per PMU

BPA Synchrophasor Lab



- Flat binary files
- 1 minute per file
- BPA .pdat format

Why Data Mining?

- BPA lab archive: 3 TB per month, up to 200 TB
  - Data from more than 140 PMUs
- Researchers developing SP-based algorithms that require long-term validation
  - Verify operation, tune parameters, avoid false positives
- Alarm limits established with long-term baselining
  - Ex: “Normal” vs. “Abnormal” phase angle
- Next gen event detection uses machine learning

Synchrophasor Data Mining

- More than event detection in post-real-time
- Extracting useful information from large quantities of stored data
- Data may be in historian/database or flat files
- Varying levels of availability and “quality”
- Mining tools must be flexible, robust, limit vulnerability to false positives
- Comprehensive approach including multiple data sources

Data Formats

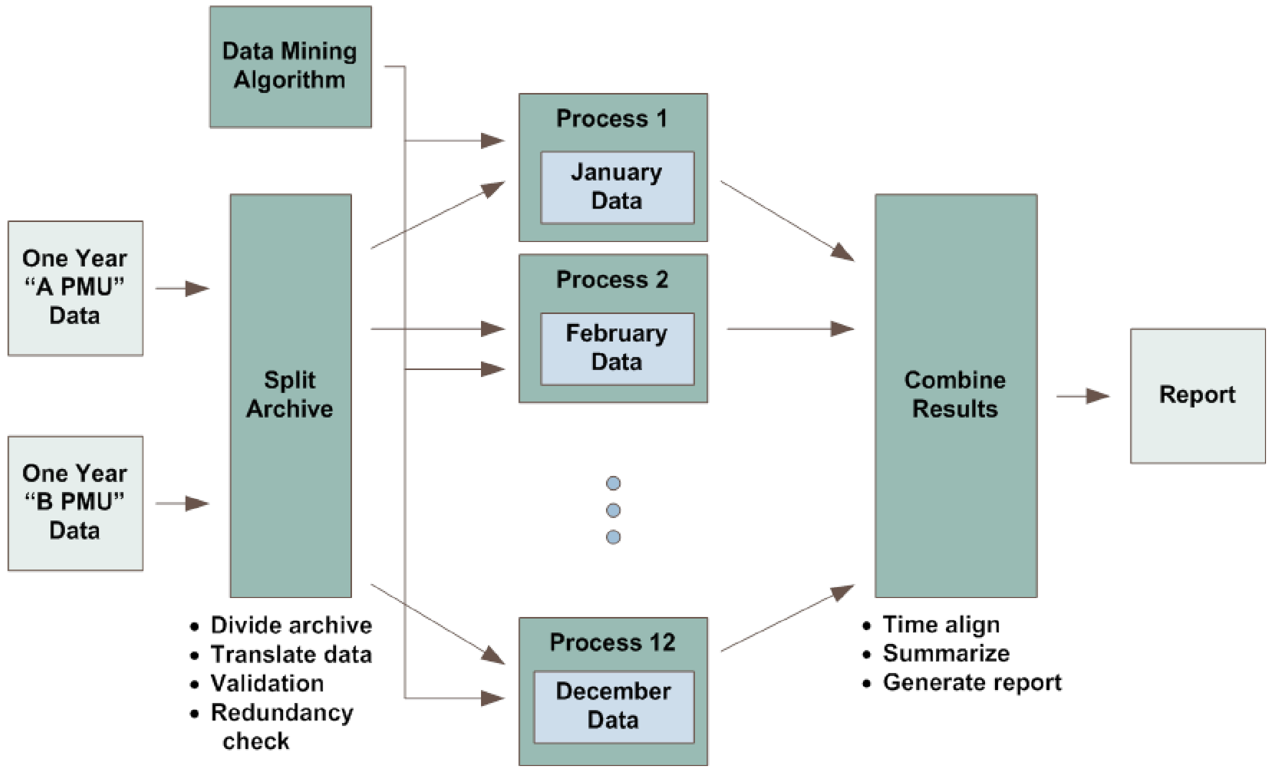
- Real Time/Streaming – C37.118, IEC 61850
- Binary Archive – BPA .pdat, COMTRADE
- ASCII Archive - .csv, COMTRADE
- Historian/Database – PI, OpenHistorian
- Supplemental Data – SCADA, state estimator relay/DFR records, SER, weather, etc.
- Metadata – System models, measurement locations, naming conventions, etc.

Potential Positives/Approaches

- Highly linear, time-sequenced
  - Ideal for parallel processing, clustering
- Abnormal values can be indexed – 99+% "normal"
  - Min and Max for day, hour, minute
- Scaled compression/downsampling
  - X months full fidelity, Y months compressed
- Possible spatial reduction based on redundancy, similar measurements

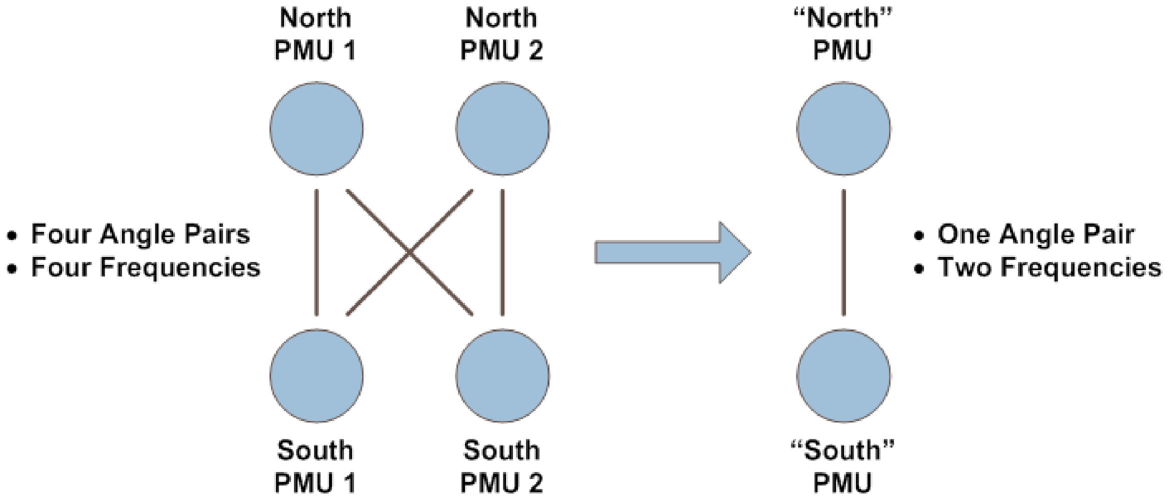
Parallel Processing Example

- Distribute archive among multiple processes



Spatial Reduction Example

- Combine measurements from related locations to reduce calculation



Potential Negatives/Challenges

- Hardware limitations – storage, processing
- No standard definition of “event”
  - Threshold, rate of change, oscillations, etc.
- Access to multiple data sources
  - Data silos, separate networks
- Changing data sets over time
  - New PMUs, changing signals/names, etc.
- Visualization, reporting of results

Potential Negatives/Challenges

- Bad or missing data
  - Similar characteristics as events
- No widely-used standard archive format
  - Data retrieval with proprietary formats
  - Algorithm exchange difficult
  - Researcher expectations vs. utility development
- Conversion/translation may be necessary
  - Floating point/integer, polar/rectangular
- Time alignment from multiple data sources
  - Ex: SCADA vs. PMU

Contact

Tony Faris
Bonneville Power Administration
Measurement Systems
(360) 418-2005
ajfaris@bpa.gov